ABSTRACT OF THE DISCLOSURE

Disclosed is a method, a computer program product and a system for processing documents that contain chemical names. The system has a unit to partition document text and to assign semantic meaning to words; a unit to recognize any substructures present in the chemical name fragments; and a unit to determine structural connectivity information of the chemical name fragments and recognized substructures and to store the determined structural connectivity information in a searchable index. The system further includes a unit to search a text index using at least one of a fragment name and a substructure name and to search the structure index by at least one of fragment connectivity and substructure connectivity. At an intersection of the search results from the structure index and the text index, the system operates to identify at least one document that contains a reference to a corresponding chemical compound.